Order-2 Asymptotically Optimal Decentralized Detection with Discrete-time Observations and 1-bit Messaging

Yasin Yilmaz[†] and Xiaodong Wang[†]

Abstract

We consider decentralized detection for wireless sensor networks. A sequential scheme based on level-triggered sampling is proposed. In the proposed scheme, sensors compute log-likelihood ratio (LLR) of their local observations, sample local LLR using level-triggered sampling and transmit a single bit at each sampling time to a fusion center (FC). At each sampling time excess LLR over (below) sampling threshold is linearly encoded in time. The FC, upon receiving a bit from a sensor, decodes excess LLR and updates approximate global LLR, which it uses as test statistic. An SPRT-like test is used by the FC to reach a final decision. We show that the proposed scheme achieves order-2 asymptotic optimality by using only a single bit for each sample, thanks to time-encoding the overshoot.

I. Introduction

We consider the problem of binary decentralized detection where a number of distributed sensors, under bandwidth constraints, communicate with a fusion center (FC) which is responsible for making the final decision. Most works on decentralized detection, e.g., [1], [2], treat the fixed-sample-size approach where each sensor collects a fixed number of samples and the FC makes its final decision at a fixed time. There is also a significant volume of literature that considers the sequential detection approach, e.g., [3]–[5]. Data fusion (multi-bit messaging) is known to be much more powerful than decision fusion (one-bit messaging) [6], albeit it consumes higher bandwith. Moreover, the recently proposed sequential detection schemes based on level-triggered sampling in [5] and [7] are as powerful as data-fusion techniques, and at the same time they are as simple and bandwidth-efficient as decision-fusion techniques.

[†]Electrical Engineering Department, Columbia University, New York, NY 10027.

Sequential probability ratio test (SPRT) is the optimum sequential detector in terms of minimizing the average detection delay for independent and identically distributed (i.i.d.) observations [**?**]. Moreover, it significantly outperforms its fixed-sample-size counterparts in minimizing the average detection delay. For instance, it has been shown in [9, Page 109] that for Gaussian signals the average detection delay (sample number) of SPRT is asymptotically one-fourth of that of the best fixed-sample-size detector satisfying the same error probability constraints. We will next formulate the problem of interest and explain SPRT as the optimal centralized sequential detector.

II. Problem Formulation and SPRT

Consider a wireless sensor network with K sensors and a fusion center (FC) which is responsible for making the final decision. Each sensor observes a discrete-time signal $y_t^k,\ t \in \mathbb{N},\ k = 1, \dots, K$, which is completely available to the FC in the centralized detection problem. In such a centralized system, the FC, using the local observations $\{y_t^k\}_{t,k}$ [1], computes a test statistic and performs a sequential test to select between two hypotheses: the signal of interest is present (H_1) or not (H_0). Specifically, in a sequential test the decision maker, i.e., FC, at each time t either makes its final decision (H_0 or H_1) or continues to receive observations, unlike in a fixed-sample-size test where the FC waits until a certain time and decides between H_0 and H_1 at that specific time.

The binary hypothesis testing problem that we consider at time t can be formulated as follows [2],

$$\{y_\tau^k\}_{\tau,k} \sim \begin{cases} f_0, & \mathsf{H}_0 \\ f_1, & \mathsf{H}_1 \end{cases} \tag{1}$$

where $\sim$ denotes "distributed according to" and $f_i,\ i = 0, 1$ is the joint probability density function (pdf) of the observed signals under H_i. SPRT at time t tests H_1 against H_0 by computing the log-likelihood ratio (LLR) as

$$L_t \triangleq \log \frac{f_1(\{y_\tau^k\}_{\tau,k})}{f_0(\{y_\tau^k\}_{\tau,k})}, \tag{2}$$

and then comparing it to some thresholds A and $-B$. Specifically, the FC at each time t makes the following decision

$$\delta_t \triangleq \begin{cases} \mathsf{H}_0 & \text{if } L_t \leq -B, \\ \mathsf{H}_1 & \text{if } L_t \geq A, \\ \text{continue} & \text{if } L_t \in (-B, A). \end{cases} \tag{3}$$

[1]The subscripts t and k denote $t \in \mathbb{N}$ and $k = 1, \dots, K$, respectively.

[2]The subscript τ denotes $\tau = 1, \dots, t$.

In other words, the FC either waits until time $t+1$ to see the new observations $\{y_{t+1}^k\}_k$ if $-B < L_t < A$, or terminates the scheme otherwise. If the scheme is terminated at time t, then the FC decides on either H_0 if $L_t \leq -B$, or H_1 if $L_t \geq A$. The thresholds $(A, B > 0)$ are selected so that the error probability constraints $\mathsf{P}_0(\delta_{\mathcal{T}} = \mathsf{H}_1) \leq \alpha$ and $\mathsf{P}_1(\delta_{\mathcal{T}} = \mathsf{H}_0) \leq \beta$ are satisfied with equalities [3], where α, β are target error probability bounds, and

$$\mathcal{T} \triangleq \min\{t \in \mathbb{N} : L_t \notin (-B, A)\} \tag{4}$$

is the detection delay (decision delay). The remarkable property of SPRT is that it simultaneously minimizes $\mathsf{E}_0[\mathcal{T}]$ and $\mathsf{E}_1[\mathcal{T}]$ while satisfying the error probability constraints [4].

Assuming independence among the observations of different sensors we can write the LLR in (2) as

$$L_t = \sum_{k=1}^{K} L_t^k, \quad L_t^k \triangleq \log \frac{f_1^k(\{y_\tau^k\}_\tau)}{f_0^k(\{y_\tau^k\}_\tau)}, \tag{5}$$

where $f_i^k,\ i = 0, 1$ denotes the joint pdf of the signal observed by sensor k under H_i. Further assume that the observations at the same sensor are i.i.d.. Then, the local LLR L_t^k is given by

$$L_t^k = \sum_{\tau=1}^{t} l_\tau^k, \quad l_\tau^k \triangleq \log \frac{f_1^k(y_\tau^k)}{f_0^k(y_\tau^k)}, \tag{6}$$

where l_τ^k is the LLR of the single observation y_τ^k.

III. Decentralized Detection via Level-triggered Sampling

In the centralized detection problem, considered in the previous section, it is assumed that each sensor k sends its local LLR L_t^k, given in (5), to the FC at each time t with infinite number of bits. In fact, it suffices to send l_t^k, given in (6), at each time t. On the other hand, in a decentralized system this is not possible due to some stringent power and bandwith constraints on sensors. In particular, each sensor k needs to sample its local LLR signal L_t^k and quantize the sampled values to comply with the power and bandwith constraints, respectively. Furthermore, the number of quantization bits per sample should be considerably low, e.g., 1 bit. A trivial solution to this decentralized problem is to use first the conventional uniform-in-time sampling, and then a simple quantizer (e.g., mid-tread, mid-riser). It has been recently shown in [5], [7] that a smarter choice is to use the *level-triggered* sampling, which is a simple and popular member of the family of *event-triggered* sampling techniques. Level-triggered

[3] $\mathsf{P}_i,\ i = 0, 1$ denotes the probability measure under H_i. Note that at the stopping time $\mathcal{T}$ the decision $\delta_{\mathcal{T}}$ is either H_0 or H_1, i.e., the thresholds A, B are selected so that $\mathsf{P}_0(\delta_{\mathcal{T}} = \mathsf{H}_0) = 1 - \alpha$ and $\mathsf{P}_1(\delta_{\mathcal{T}} = \mathsf{H}_1) = 1 - \beta$.

[4] $\mathsf{E}_i,\ i = 0, 1$ denotes the expectation under H_i.

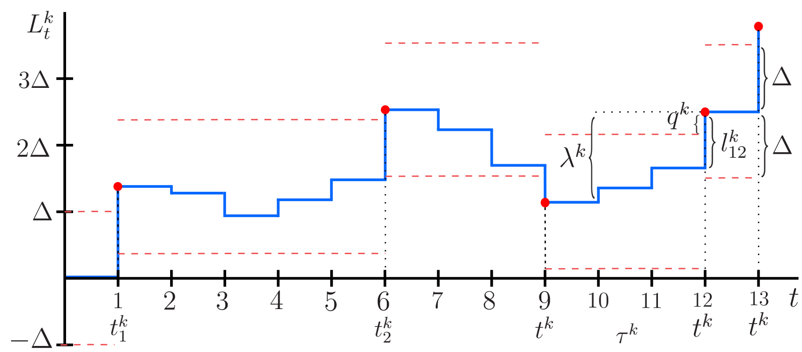

Fig. 1. The level-triggered sampling mechanism with sampling levels $\Delta, -\Delta$ is shown on a sample path of the local LLR signal L_t^k.

sampling, producing 1-bit information per sample, eliminates the need for a quantization step, which is clearly a big advantage for power and bandwith-limited systems. Then, at the FC side the decentralized detection schemes based on level-triggered sampling, given in [5], [7], imitating the optimal sequential detector approximate the LLR L_t and apply an SPRT-like test [cf. (3), (4)] using this approximation as the test statistic.

A. Level-triggered Sampling

In level-triggered sampling, the local LLR signal L_t^k is sampled at a sequence of random times $\{t_n^k\}_n$ [5], as shown in Fig. 1. On the contrary, in the conventional uniform sampling, sampling times are deterministic and uniform in time, e.g., $\{mT : m \in \mathbb{N}^+\}$ with a sampling interval T. The sampling intervals in level-triggered sampling, $\{\tau_n^k : \tau_n^k \triangleq t_n^k - t_{n-1}^k\}_n,\ t_0^k = 0$, are also random. The sampling times $\{t_n^k\}_n$ are dynamically determined by the signal to be sampled L_t^k. In other words, $\{t_n^k\}_n$ depend on the realizations of the random signal L_t^k, hence they are random. Specifically, as shown in Fig. 1 the nth sample is taken at time t_n^k when the change in the signal since the last sampling time t_{n-1}^k exceeds either Δ or $-\Delta$, i.e.,

$$t_n^k \triangleq \min\{t \in \mathbb{N} : |L_t^k - L_{t_{n-1}^k}^k| \geq \Delta\},\ L_0^k = 0, \tag{7}$$

where $\Delta > 0$ is a predetermined level.

Note the similarity between the formulations of the decision time in (4), which is the stopping time of SPRT, and the sampling time in (7). In fact, the level-triggered sampling procedure is a succession of identical SPRTs whose stopping times are given in (7). The thresholds of this local SPRT are $\Delta, -\Delta$ and it is repeated after each sampling time t_n^k until the FC terminates the scheme. For instance, the nth

[5]The subscript n denotes $n \in \mathbb{N}^+$

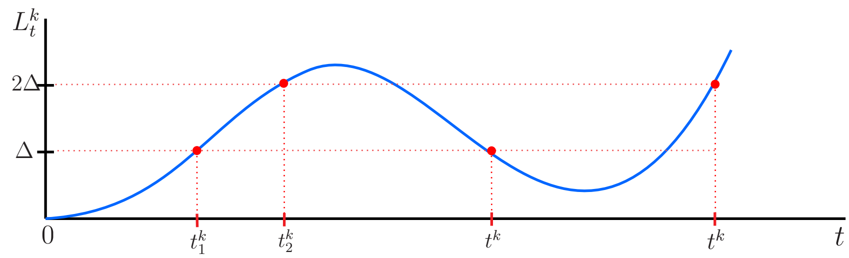

Fig. 2. The level-triggered sampling mechanism with sampling levels $\Delta, -\Delta$ is shown for the case where sensors observe continuous-time signals with continuous paths, discussed in [5].

local SPRT starts at time $t_{n-1}^k + 1$, stops at the nth sampling time t_n^k and the nth sampling interval τ_n^k represents the number of observations used in the nth local SPRT. Denote the stopping LLR value of the nth local SPRT with λ_n^k, i.e., $\lambda_n^k \triangleq L_{t_n^k}^k - L_{t_{n-1}^k}^k = \sum_{t=t_{n-1}^k+1}^{t_n^k} l_t^k$ (cf. Fig. 1). Then, the decision of the nth local SPRT can be written as

$$b_n^k \triangleq \text{sign}(\lambda_n^k), \tag{8}$$

where from (7) we have either $\lambda_n^k \geq \Delta$ or $\lambda_n^k \leq -\Delta$. These local SPRTs also induce local error probabilities, however there is no constraint on such local error probabilities since local SPRTs, i.e., level-triggered sampling, are used only as a mean of transmitting information from sensors to the FC. Therefore, we are in a sense free to set the local threshold Δ. The choice of Δ obviously affects the sampling intervals $\{\tau_n^k\}_{n,k}$, hence it can be set to ensure a specific value for the average sampling interval $\mathsf{E}_i[\tau_n^k]$. It has been shown in [7, Section IV-A] that the equation $\Delta \tanh(\frac{\Delta}{2}) = T \frac{\sum_{k=1}^K \mathsf{E}_i[l_t^k]}{K}$ can be used to determine Δ for the average sampling interval value T. This important duality result between level-triggered sampling and SPRT will let us use the properties of SPRT to analyze the level-triggered sampling procedure.

The sample summary (local decision) b_n^k, given in (8), is a 1-bit encoding of the change λ_n^k in the LLR signal L_t^k during the time interval $(t_{n-1}^k, t_n^k]$. Put another way, b_n^k represents the threshold (Δ or $-\Delta$) exceeded by λ_n^k. However, it does not represent how much λ_n^k exceeded the threshold. Define $q_n^k \triangleq |\lambda_n^k| - \Delta$ as the excess amount of LLR over Δ or under $-\Delta$, namely overshoot [cf. Fig. 1]. Then, λ_n^k is given by

$$\lambda_n^k = b_n^k(\Delta + q_n^k). \tag{9}$$

B. Decentralized Detection Scheme

In our decentralized detection scheme, each sensor k transmits b_n^k to the FC at time $s_n^k \in [t_n^k, t_n^k + 1)$. Assume for now that the FC, knowing t_n^k and s_n^k, somehow computes q_n^k. We will describe how it

computes in the following subsection. Then, upon receiving b_n^k at time s_n^k it recovers λ_n^k by using (9), and computes $L_{t_n^k}^k$ as follows

$$L_{t_k^n}^k = \sum_{m=1}^{n} L_{t_m^k}^k - L_{t_{m-1}^k}^k = \sum_{m=1}^{n} \lambda_m^k. \tag{10}$$

In other words, the local LLR $L_{t_n^k}^k$ is available to the FC at time s_n^k with a delay of $\xi_n^k \triangleq s_n^k - t_n^k \in [0,1), \forall k, n$.

In this case where the FC can compute the overshoot q_n^k and accordingly the LLR sample $L_{t_n^k}^k$, there is loss, other than the small bounded delay ξ_n^k, only in the time resolution (but not in the magnitude resolution) while recovering the local LLR signal L_t^k. Similarly, in the case where sensors observe and process continuous-time signals with continuous paths, e.g., Wiener process, the only loss is in the time resolution [5]. The LLR L_t^k of such a continuous-time signal y_t^k is also a continuous-time signal, e.g., the one given in Fig. 2. In this continuous-time case, as can be seen in Fig. 2, L_t^k hits the sampling levels exactly, hence no overshoot occurs, i.e., $q_n^k = 0, \forall n, k$. On the other hand, in a discrete-time case where the overshoots $\{q_n^k\}_{n,k}$ are not available to the FC, there is loss also in the magnitude resolution. To overcome this overshoot problem [5] and [7] followed different approaches. In [5], the LLR of b_n^k is computed via simulations, which in fact includes an average value for q_n^k. At each sampling time t_n^k this average value is used to replace the unknown overshoot q_n^k in (9). However, this approach brought about an asymptotic optimality of order-1, but not order-2 which is a stronger type of asymptotic optimality [6], achieved in the continuous-time case. In [7], at each time t_n^k some additional bits are used to quantize the overshoot q_n^k. Following this approach it has been shown that order-2 asymptotic optimality can be attained by increasing the number of quantization bits at a rate of $\log|\log\gamma|$ where $\gamma \to 0$ at least as fast as the error probabilities α and β, i.e., $\gamma = O(\alpha), \gamma = O(\beta)$.

Returning back to our decentralized detection scheme based on level-triggered sampling, the FC, using (10), approximates the local and global LLR signals L_t^k and L_t with the stepwise functions $\hat{L}_t^k$ and $\hat{L}_t$, respectively as follows

$$\begin{aligned} \hat{L}_t^k =& L_{t_n^k}^k, \quad t \in [s_n^k, s_{n+1}^k), \quad t_0^k = 0, \\ \text{and} \quad \hat{L}_t =& \sum_{k=1}^{K} \hat{L}_t^k. \end{aligned} \tag{11}$$

In fact, the FC employs the following procedure. Upon receiving the nth bit b_n from any sensor at time s_n (with a short delay ξ_n after b_n is actually sampled at time t_n) it computes λ_n as in (9) (assuming the

[6]The definitions of asymptotic optimality types are given in Section IV-A.

overshoot q_n is computable by knowing t_n, s_n) and updates the approximate global LLR $\hat{L}_t$ as

$$\hat{L}_{s_n} = \hat{L}_{s_{n-1}} + \lambda_n. \tag{12}$$

The approximate LLR signal $\hat{L}_t$ is kept constant until the arrival of the next bit b_{n+1} from any sensor unless the scheme terminates at time s_n. Following SPRT, given in (3), (4), an SPRT-like procedure is applied at the FC to terminate the scheme. Specifically, the scheme is terminated at the stopping time $\hat{\mathcal{T}}$ given by

$$\hat{\mathcal{T}} \triangleq \min\{t \in \mathbb{N} : \hat{L}_t \notin (-\hat{B}, \hat{A})\}, \tag{13}$$

where the thresholds $(\hat{A}, \hat{B} > 0)$ are selected to satisfy the error probability constraints $\mathsf{P}_0(\hat{\delta}_{\hat{\mathcal{T}}} = \mathsf{H}_1) \leq \alpha$ and $\mathsf{P}_1(\hat{\delta}_{\hat{\mathcal{T}}} = \mathsf{H}_0) \leq \beta$ with equality. The decision function $\hat{\delta}_{\hat{\mathcal{T}}}$ is written as

$$\hat{\delta}_{\hat{\mathcal{T}}} \triangleq \begin{cases} \mathsf{H}_0 & \text{if } \hat{L}_{\hat{\mathcal{T}}} \leq -\hat{B}, \\ \mathsf{H}_1 & \text{if } \hat{L}_{\hat{\mathcal{T}}} \geq \hat{A}. \end{cases} \tag{14}$$

A "stop" signal is sent to all sensors immediately after the global decision $\hat{\delta}_{\hat{\mathcal{T}}}$ is made at the FC.

Note from (11) that although the approximate local LLR $\hat{L}_t^k$ attains the value of the exact local LLR $L_{t_n^k}^k$ at time s_n^k, shortly after the sampling time t_n^k, i.e., $\hat{L}_{s_n^k}^k = L_{t_n^k}^k$, this is in general not true for the approximate global LLR $\hat{L}_t$, i.e., $\hat{L}_{s_n^k} \neq L_{t_n^k}$, since $\hat{L}_{s_n^k}^j \neq L_{t_n^k}^j,\ j \neq k$, for the other $k-1$ sensors. Next, we explain how the FC computes the overshoot q_n^k by knowing t_n^k and s_n^k.

C. *Time-encoding the Overshoot*

Overshoot q_n^k, limiting the performance of the decentralized detection schemes based on level-triggered sampling, poses an important problem. As shown in Fig. 1 it is a random variable occurring at each sampling time t_n^k and cannot be made available to the FC by sending the sign bit b_n^k, given in (8), at time t_n^k. In other words, q_n^k represents the missing LLR information at the FC, which is not encoded in b_n^k, to recover the local LLR information λ_n^k accumulated during the interval $(t_{n-1}^k, t_n^k]$. We have to tackle this overshoot problem for good asymptotic performance since ignoring it causes such missing LLR information, i.e., overshoot error, to unboundedly accumulate in recovering the local LLR signal L_t^k, which in turn results in poor asymptotic performance. Using an average value in lieu of the random overshoot q_n^k in each sample helps partially, enabling us to achieve an order-1 asymptotically optimal average detection delay performance [5]. Quantizing the overshoot by using extra bits (in addition to the sign bit) ensures order-2 asymptotic optimality, but requires infinite number of bits asymptotically [7]. Thus, it is not a perfect solution either.

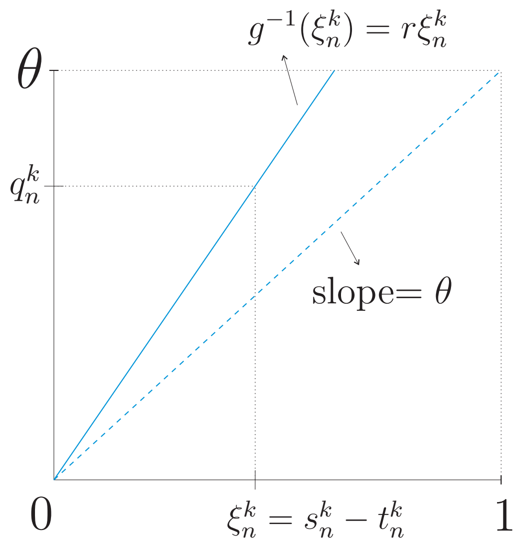

Fig. 3. The linear decoding function g^{-1} with slope $r > \theta$ maps the transmission delay ξ_n^k to the overshoot $q_n^k = r\xi_n^k$.

In this paper, we propose to linearly encode the overshoot in time, i.e., not in extra bits, hence requiring only one bit, namely the sign bit b_n^k, given in (8). We assume a linear function $g(q_n^k) = \frac{q_n^k}{r}$ to encode the overshoot q_n^k in each sample. The slope parameter r of the linear function is known to sensors and the FC. The output of the encoding function is the transmission delay $\xi_n^k = s_n^k - t_n^k$, that is the delay between the transmission time s_n^k and the sampling time t_n^k. Assuming a global clock, i.e., $t \in \mathbb{N}$ is the same for all sensors and the FC, the FC upon receiving b_n^k at time s_n^k can compute ξ_n^k, provided that $\xi_n^k \in [0, 1), \forall k, n$. Then, the overshoot can be recovered by using the inverse encoding function $g^{-1}(\xi_n^k)$, i.e., $q_n^k = r\xi_n^k$, as shown in Fig. 3.

To this end, we made the following assumptions.

(A1) There exists a global clock t, running in the wireless sensor network, hence the FC knows the potential sampling times *a priori*.

(A2) Each transmission delay is bounded by unit time, i.e., $\xi_n^k \in [0, 1), \forall k, n$. We can easily justify this assumption by selecting the slope parameter r as $r > q_n^k, \forall k, n$. In other words, this assumption is equivalent to assuming a bound θ for overshoots and setting $r > \theta = \max_{k,n} q_n^k$.

(A3) We have ideal channels with no delay or deterministic delays between sensors and the FC. That is, each pair (s_n^k, b_n^k) of transmission time and transmitted bit is available to the FC.

Having (A1), (A2) and (A3) each sampling time t_n^k can be uniquely determined at the FC with the reception of each b_n^k (*a posteriori*). If (A1) is violated, i.e., the network lacks a global clock, then a synchronization bit (in addition to b_n^k) can be used to report each sampling time t_n^k, resulting in a 2-

bit scheme. Note that (A2) defines a lower bound for r. The slope of the inverse encoding (decoding) function g^{-1} must be larger than the maximal overshoot value θ, as shown in Fig. 3. We do not have an upper bound for r, and in fact by choosing r arbitrarily large we can have each transmission delay ξ_n^k arbitrarily small. If channel delays are random, i.e., (A3) is violated, then the transmission delay ξ_n^k, and thus the overshoot q_n^k, in general, are not exactly recovered at the FC. However, if the FC estimates this random channel delay well enough, then the scheme still achieves a good performance, as will be analyzed in Section IV-B.

IV. Performance Analysis

In this section, we will analyze the asymptotic average detection delay performance of the scheme proposed in the previous section, and also the effect of random channel delay on the proposed scheme.

A. Asymptotic Analysis on Average Detection Delay

We use the following definition for asymptotic optimality.

Definition. *A sequential scheme $(\hat{\mathcal{T}}, \hat{\delta}_{\hat{\mathcal{T}}})$, with stopping time $\hat{\mathcal{T}}$ and decision function $\hat{\delta}_{\hat{\mathcal{T}}}$, satisfying the two error probability constraints $\mathsf{P}_0(\hat{\delta}_{\hat{\mathcal{T}}} = 1) \leq \alpha$ and $\mathsf{P}_1(\hat{\delta}_{\hat{\mathcal{T}}} = 0) \leq \beta$ is said to be order-1 asymptotically optimal if*

$$\frac{\mathsf{E}_i[\hat{\mathcal{T}}]}{\mathsf{E}_i[\mathcal{T}]} = 1 + o(1) \quad \textit{as } \alpha, \beta \to 0, \tag{15}$$

where $\mathcal{T}$ is the stopping time of the optimal SPRT that satisfies the error probability constraints with equality and $o(1)$ denotes a vanishing term as $\alpha, \beta \to 0$. It is order-2 asymptotically optimal if

$$\mathsf{E}_i[\hat{\mathcal{T}}] - \mathsf{E}_i[\mathcal{T}] = O(1), \tag{16}$$

where $O(1)$ denotes a constant term, and order-3 asymptotically optimal if

$$\mathsf{E}_i[\hat{\mathcal{T}}] - \mathsf{E}_i[\mathcal{T}] = o(1). \tag{17}$$

Note that as the order increases we have a stronger type of asymptotic optimality, i.e., order-3 $\Rightarrow$ order-2 $\Rightarrow$ order-1, where $\Rightarrow$ means "implies". Order-1 asymptotic optimality is the most commonly used type in the literature, but also the weakest type. The average detection delay of an order-1 scheme can diverge from that of the optimal scheme. On the other hand, for order-2 and order-3 schemes the average detection delay remains parallel and converges to that of the optimal scheme, respectively. Order-3 asymptotic optimality is extremely rare in the Sequential Analysis literature, and corresponds to schemes that are considered as optimum *per se* for all practical purposes.

In the following theorem, we show order-2 asymptotic optimality for the proposed scheme.

Theorem 1. *The decentralized detection scheme proposed in Section III-B is order-2 asymptotically optimal in terms of average detection delay, i.e.,*

$$\mathsf{E}_i[\hat{\mathcal{T}}] - \mathsf{E}_i[\mathcal{T}] = O(1),\ i = 0, 1, \tag{18}$$

where the stopping times $\mathcal{T}$ and $\hat{\mathcal{T}}$ are given in (4) *and* (13), *respectively.*

Proof: It was shown in [10, Theorem 2] that the asymptotic expressions for average detection delays $\mathsf{E}_i[\hat{\mathcal{T}}], i = 0, 1$ of a decentralized detection scheme based on level-triggered sampling, e.g., the one we presented in Section III, are in general written as

$$\mathsf{E}_1[\hat{\mathcal{T}}] = \frac{|\log\alpha|}{\hat{I}_1(1)} + O(1) \quad \text{and} \quad \mathsf{E}_0[\hat{\mathcal{T}}] = \frac{|\log\beta|}{\hat{I}_0(1)} + O(1), \tag{19}$$

as $\alpha, \beta \to 0$, where $\hat{I}_i(1) = \sum_{k=1}^K \frac{\hat{I}_i^k(t_1^k)}{I_i^k(t_1^k)} I_i^k(1), i = 0, 1$ are hypothetical Kullback-Leibler (KL) information numbers defined to serve analytical purposes, $I_i^k(1) = |\mathsf{E}_i[l_1^k]|$, $I_i^k(t_1^k) = |\mathsf{E}_i[\lambda_1^k]|$, $\hat{I}_i^k(t_1^k) = |\mathsf{E}_i[\hat{\lambda}_1^k]|, i = 0, 1$ are the KL information numbers of the LLR sequences $\{l_t^k\}_t$, $\{\lambda_n^k\}_n$, $\{\hat{\lambda}_n^k\}_n$, respectively. Here, at sensor k and under H_i, $\mathsf{E}_i[l_1^k]$ is the average LLR of a single observation [cf. (6)], $\mathsf{E}_i[\lambda_1^k]$ is the average LLR accumulated in a local SPRT (i.e., during a sampling interval) [cf. (8)], and $\mathsf{E}_i[\hat{\lambda}_1^k]$ is the average LLR transmitted in a bit b_n^k. Hence, the hypothetical KL information number $\hat{I}_i(1)$ is, in fact, the projection of $I_i(1) = \sum_{k=1}^K I_i^k(1)$, which is the KL information number of the global LLR sequence $\{\sum_{k=1}^K l_t^k\}$, onto the filtration generated by the transmitted bit sequence $\{b_n^k\}$. Since sensors do not transmit the LLR of each single observation (they transmit LLR messages of several observations), $\hat{I}_i(1)$ is not a real KL information number, but it is a projected or hypothetical KL information number, which plays a key role in the asymptotic analysis on average detection delay [10]. If the random overshoot q_n^k is not fully encoded in each transmitted bit b_n^k, then this projected KL information number $\hat{I}_i(1)$, as in [5], [7], [10], will not be the same as the original KL information number $I_i(1)$ of the global LLR sequence $\{\sum_{k=1}^K l_t^k\}$ since the average LLR in a transmitted bit will be different from the average LLR accumulated in a local SPRT, i.e., $\hat{I}_i^k(t_1^k) \neq I_i^k(t_1^k)$. However, for our proposed scheme we will have $\hat{I}_i^k(t_1^k) = I_i^k(t_1^k)$ since we encode the overshoot q_n^k into the transmission delay ξ_n^k, as explained in Section III-C. Therefore, from (19) we write the asymptotic average detection delays of the proposed scheme as

$$\mathsf{E}_1[\hat{\mathcal{T}}] = \frac{|\log\alpha|}{I_1(1)} + O(1) \quad \text{and} \quad \mathsf{E}_0[\hat{\mathcal{T}}] = \frac{|\log\beta|}{I_0(1)} + O(1), \tag{20}$$

as $\alpha, \beta \to 0$.

For the optimal centralized scheme we have, from [7, Lemma 2], $\mathsf{E}_1[\mathcal{T}] \geq \frac{|\log\alpha|}{I_1(1)} + o(1)$ and $\mathsf{E}_0[\mathcal{T}] \geq$

$\frac{|\log\beta|}{I_0(1)} + o(1)$, which is, together with (20), sufficient to prove order-2 asymptotic optimality, i.e, (18). In order to show that we cannot achieve order-3 asymptotic optimality, following [10, Theorem 2] we can write $\mathsf{E}_1[\mathcal{T}] = \frac{|\log\alpha|}{I_1(1)} + O(1)$ and $\mathsf{E}_0[\mathcal{T}] = \frac{|\log\beta|}{I_0(1)} + O(1)$, where the $O(1)$ term includes the average excess LLR in $L_{\mathcal{T}}$ over A or below $-B$. Note that the excess LLR here is produced by the global LLR $\sum_{k=1}^{K} l_{\mathcal{T}}^k$ of the single observations at the stopping time $\mathcal{T}$ (cf. Section II). On the other hand, the average excess LLR included in the $O(1)$ term in (20) is a consequence of the final LLR message(s) received at the stopping time $\hat{\mathcal{T}}$ from any sensor(s) (cf. Section III-B). Obviously, the average excess LLR produced by message(s), including LLR of several observations, is larger than that produced by single observations. Hence, the difference $\mathsf{E}_i[\hat{\mathcal{T}}] - \mathsf{E}_i[\mathcal{T}]$ remains as a bounded constant. ■

Note that by using only a single bit we achieve order-2 asymptotic optimality in the discrete-time case, similar to the continuous-time case as analyzed in [5]. This is a remarkable performance improvement for decentralized detection schemes. It was shown in [7] that level-triggered sampling-based scheme with a single bit, i.e., disregarding the overshoot problem, significantly outperforms the scheme based on conventional uniform-in-time sampling with a single bit. Moreover, it even outperforms the uniform sampling-based scheme with infinite number of bits. Now, the proposed scheme significantly outperforms the previously introduced level-triggered sampling-based schemes by achieving order-2 asymptotic optimality with only a single bit. As noted in the previous section, with discrete-time observations the scheme in [5] achieves order-1 asymptotic optimality using a single bit, and the scheme in [7] achieves order-2 asymptotic optimality using multiple bits. The number of bits required in [7] for order-2 asymptotic optimality increases with a reasonably slow rate [i.e., $\log|\log\gamma|$ where $\gamma = O(\alpha), \gamma = O(\beta)$] to infinity. For practical purposes it suffices for this scheme to use at most 5 bits to attain an order-2-like performance. In this paper, the proposed scheme, eliminating the foremost overshoot problem, with a single bit attains order-2 asymptotic optimality in theory (not only for practical purposes!). The asymptotic optimality results in the above references and also this paper assume ideal channels between sensors and the FC, i.e., transmitted bits are available to the FC with either no delay or deterministic delays. We next analyze how the performance of the proposed scheme is affected by random channel delays.

B. Random Delay in Bit Arrival Times

Consider a random channel delay ν_n^k in the arrival time of each transmitted bit b_n^k. In other words each b_n^k, transmitted from sensor k at time s_n^k, is received by the FC at time $s_n^k + \nu_n^k$ where ν_n^k is random and $\nu_n^k \geq 0$. Hence, the transmission delay together with the random channel delay, $\xi_n^k + \nu_n^k$, gives us the reception delay. We assume that the channel delay ν_n^k is bounded by a constant ϕ, i.e., $0 \leq \nu_n^k \leq \phi, \forall k, n$.

To counteract this nonnegative channel delay we estimate it with $\hat{\nu}_n^k$. Subtracting $\hat{\nu}_n^k$ from each reception delay we obtain an estimate $\hat{\xi}_n^k = \xi_n^k + \nu_n^k - \hat{\nu}_n^k$ for each transmission delay ξ_n^k. Note that the estimation error $|\hat{\xi}_n^k - \xi_n^k|$ is bounded by a constant $\hat{\phi}$, i.e.,

$$|\hat{\xi}_n^k - \xi_n^k| = |\hat{\nu}_n^k - \nu_n^k| \leq \hat{\phi},\ \forall k, n. \tag{21}$$

Even with a simple-minded estimator that uses $\mathsf{E}[\nu_n^k]$ as its constant estimate for each ν_n^k, we have $\hat{\phi} < \phi$. Using a better estimator we can have a bound $\hat{\phi} \ll \phi$.

Finally, using each $\hat{\xi}_n^k$ in the decoding function g^{-1} we compute an estimate $\hat{q}_n^k$ for each overshoot q_n^k, i.e., $\hat{q}_n^k = g^{-1}(\hat{\xi}_n^k) = r\hat{\xi}_n^k$. The overshoot estimation error $|\hat{q}_n^k - q_n^k|$ is then given by

$$|\hat{q}_n^k - q_n^k| \leq r\hat{\phi},\ \forall k, n. \tag{22}$$

From (22), we see that the overshoot estimation error depends on the accuracy of the delay estimation and also the choice of the slope of the decoding function. Note that larger the slope r is, larger the overshoot estimation error will be. Recall the advantage of having large r stated in Section III-C. Through the encoding function g, large r enables us to have small transmission delay ξ_n^k. This helps the FC, under ideal channels, to know each sampled local LLR value $L_{t_n^k}^k$ [cf. (11)] shortly after the corresponding sampling time t_n^k, at time $s_n^k = t_n^k + \xi_n^k$. Moreover, under ideal channels the FC can practically know each LLR sample at time t_n^k by setting r arbitrarily large. However, under non-ideal channels with random delay there is an important disadvantage of having large r, as can be seen in (22). We need to select r as small as possible to obtain better estimates of the overshoot q_n^k. Although there seems to be a trade-off in selecting r under non-ideal channels, the disadvantage of large r by far outweighs its advantage. This is because errors in overshoot estimation accumulate in time while computing the approximate LLR $\hat{L}_t$, as in (12), where now we have $\hat{\lambda}_n$ instead of λ_n. This error accumulation eventually worsens the asymptotic average detection delay performance, whereas transmission delays do not hurt the asymptotic performance since they do not accumulate, i.e., appear independently in each transmission time $s_n^k = t_n^k + \xi_n^k$, and they are bounded ($\xi_n^k < 1, \forall k, n$). Hence, in the case of random channel delays one should select r as small as possible.

Under ideal channels, we have $r > \theta$ due to (A2), as shown in Fig. 3. The assumption (A2) reads $s_n^k \in [t_n^k, t_n^k + 1), \forall k, n$. This is to ensure that the FC correctly determines the sampling time t_n^k and computes the transmission delay ξ_n^k. Here with non-ideal channels, it should read $\hat{s}_n^k \in [t_n^k, t_n^k + 1), \forall k, n$, where $\hat{s}_n^k = s_n^k + \nu_n^k - \hat{\nu}_n^k$ is the estimate of the transmission time s_n^k. From (21) we know that $\hat{s}_n^k \in [s_n^k - \hat{\phi}, s_n^k + \hat{\phi}]$, implying $s_n^k \in [t_n^k + \hat{\phi}, t_n^k + 1 - \hat{\phi})$ and thus $\xi_n^k \in [\hat{\phi}, 1 - \hat{\phi})$. Note that in this case we

have the linear encoding function $g(q_n^k) = \hat{\phi} + \frac{q_n^k}{r}$. Accordingly, $\frac{q_n^k}{r} \in [0, 1-2\hat{\phi})$ and thus $r > \frac{q_n^k}{1-2\hat{\phi}}$. Recalling that $\max_{k,n} q_n^k = \theta$ under non-ideal channels we have $r > \frac{\theta}{1-2\hat{\phi}}$. Then, selecting the smallest possible r we have

$$|\hat{q}_n^k - q_n^k| \leq \frac{\hat{\phi}}{1-2\hat{\phi}}\,\theta, \quad \forall k, n, \tag{23}$$

which follows from (22). Note from (23) that if we have a small $\hat{\phi}$, i.e., a good delay estimator, then the overshoot can be well estimated, which corresponds to a multi-bit overshoot quantization scheme as in [7]. For instance, with $\hat{\phi} = 0.02$ the proposed scheme using a single bit achieves the same performance as the scheme in [7] with 5 bits, which in practice achieves order-2 asymptotic optimality.

V. Simulations

VI. Conclusion

References

[1] J. Tsitsiklis, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals, and Systems*, pp. 167-182, 1988.

[2] P. Willett, P.F. Swaszek, and R.S. Blum, "The good, bad and ugly: distributed detection of a known signal in dependent gaussian noise," *IEEE Trans. Sig. Proc.*, vol. 48, no. 12, pp. 3266-3279, Dec. 2000.

[3] V.V. Veeravalli, T. Basar, and H.V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inform. Theory*, vol. 39, no. 2, pp. 433442, Mar. 1993.

[4] A.M. Hussain, "Multisensor distributed sequential detection," *IEEE Trans. Aero. Electron. Syst.*, vol. 30, no. 3, pp. 698-708, July 1994.

[5] G. Fellouris, and G.V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 534-548, Jan. 2011.

[6] S. Chaudhari, J. Lunden, V. Koivunen, and H.V. Poor, "Cooperative sensing with imperfect reporting channels: Hard decisions or soft decisions?," *IEEE Trans. Sig. Proc.*, vol. 60, no. 1, pp. 18-28, Jan. 2012.

[7] Y. Yilmaz, G.V. Moustakides, and X. Wang, "Cooperative sequential spectrum sensing based on level-triggered sampling," *IEEE Trans. Sig. Proc.*, vol. 60, no. 9, pp. 4509-4524, Sep. 2012.

[8] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, vol. 19, pp. 326-329, 1948.

[9] H.V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd edition, Springer, New York, NY, 1994.

[10] Y. Yilmaz, G.V. Moustakides, and X. Wang, "Channel-aware decentralized detection via level-triggered sampling," *IEEE Trans. Sig. Proc.*, vol. 61, no. 1, Jan. 2013.